PC Support Advisor

The Essential Resource for PC Support Professionals

Update 67

On This Month's Utility Disk

.DLL Manager for Windows.

Remaining Interrupt List files from PCSA 66.

Novell DOS 7 updates and patches.

"Not NerdPerfect" - says Robert Schifreen.

Everyone knew that WordPerfect was in some kind of trouble. The company had spent millions of dollars developing WordPerfect 6.0 for DOS, and there were delays of many months between the announcement and the delivery. So not only was the company spending lots of cash on developing 6, but no one was buying 5.1 until they'd had a chance to weigh it up against the new version.

When the new version finally arrived, it was buggy, slow, and required at least five times the hard disk space that 5.1 did. The only major new features were the WYSIWYG editing mode and the fax module.

Then WordPerfect made another silly mistake. It announced in public that 6 for DOS "has taken DOS word processing about as far as it's possible to go", and that there would never be a version 7. Windows, said the company, was the future, and that's where all the development effort would now be concentrated.

Unfortunately, WordPerfect for Windows is not selling in huge quantities. Not because Microsoft's Word for Windows is necessarily a better product, but simply because Microsoft can offer Word as part of a suite for around the same price as WordPerfect can get for version 6 on its own. And there's nothing that WordPerfect can do to counter this, as the company doesn't have a Windows database or spreadsheet with which to form the suite.

Quattro Pro, the spreadsheet "borrowed" from Borland, helps a little, but the way that the whole package then became marketed as "Borland Office" rather than a WordPerfect product must be one of the worst decisions that anyone at either company has made.

Then, WordPerfect finally revealed why it had admitted there would be no version 7 for DOS. It had been working on a new version of 5.1 instead, which now has a fax module and the ability to read and write version 6 files. 5.1 Plus, as it's called, is the same size and speed as 5.1, and seems like the ideal choice for DOS users who don't want to switch to version 6.

So WordPerfect 6 for DOS is no longer necessary for the majority of users, and the company prefers to promote its Windows versions so 6 for DOS is getting very little publicity.

Everyone in the industry could see that WordPerfect, as a company, was in a mess. The way it had laid off staff and started charging for support, two things that WordPerfect vowed never to do, only served to increase the speculation that the company would soon change hands.

And so, as it turns out, the industry watchers were right. The two founders of the company get around $600m each, and a seat on the board at Novell. And Microsoft is no longer the only software company that has the ability to produce applications for an operating system platform that it owns.

Both WordPerfect and Novell are trying to assure their customers that it's business as usual. Let's hope that this does indeed turn out to be the case.

PCSA

PC Support *Advisor*

PC Support *Advisor* is a publication of **International Technology Publishing**.

ISSN: 1031-3966

Editor: Robert Schifreen

Editor in Chief: Les Bell

Publisher: Chris Carvan

Email:
hex@cix.compulink.co.uk
CompuServe: 100016,301

International Technology Publishing Inc.
North American Subscriptions
Suite 200
555 De Haro Street
San Fransisco, CA94107
USA

Telephone: (415) 255 1295
Facsimile: (415) 255 8496

International Technology Publishing Inc.
European Subscriptions
13 Callcott Street
London
W8 7SU
United Kingdom

Telephone: (071) 724 9306
Facsimile: (071) 706 3296

International Technology Publishing Pty Ltd.
SE Asian Subscriptions
138 Cecil Street
Cecil Court #13-03
Singapore
0106

Telephone: (65) 227 4187
Facsimile: (65) 227 2885

International Technology Publishing Pty Ltd.
Australasian Subscriptions
Suite 511
185 Elizabeth Street
Sydney, NSW, 2000
Australia

Telephone: (02) 261 4683
Facsimile:(02) 261 4741

Understanding PC Processors

The range of processors available for today's PCs is bewildering. Peter Matthews explains what's available, and discusses the benefits of each. This article updates and replaces file H0109 which appeared in PCSA 47.

There are many points to consider when purchasing a PC. How big is the hard disk? How much RAM can you install? What's the video card like? and so on. However, of all the aspects of a particular PC, perhaps the most important one is what processor it is supplied with.

The choice of processor for your PC is now wider than it has ever been - not only is there Intel's own broad range of CPUs, many other companies are producing chips that are developed either in conjunction with Intel or entirely separately. Moreover, the design of most PCs being produced today is such that the processor can be easily replaced to allow the machine to be upgraded.

The Beginning

The story of the PC's processor really starts with the Intel 8080 chip. Although the 8080 was never used in IBM PC compatible machines, the design of it influenced the design of Intel's subsequent range of processors. The 8080 was an 8-bit processor (ie, it processed information in 8-bit chunks) with a 16-bit memory bus that gave it a maximum memory capacity of 64 KB.

Although extremely basic by today's standards, the 8080 played an important role in putting microcomputers into offices. CP/M, the operating system used by most business micros in the late 70s and early 80s, ran on the 8080. Indeed, the 8080 was so successful that other companies produced clone versions. Perhaps the most well-known of these was Zilog's Z80 processor, that could run 8080 code unchanged but added some useful facilities to the instruction set as well as being faster. The Z80 can still occasionally be found on PC expansion cards.

8088/8086

Intel realised that the 8080, though successful, wasn't good enough to compete with the new 16-bit processor (the 6800) being developed by Motorola. Therefore, it produced its own 16-bit processor called the 8086. Unlike the Motorola device, though, the 8086 had a segmented memory model whereas the 6800 had a "flat" memory space. To understand the difference, picture the machine's memory as the classic row of boxes. Each box holds a single character or instruction - a byte. With the Motorola chip, the boxes were arranged as a single long row, and numbered sequentially.

This makes memory operations simple for the programmer - moving 3000 bytes up from the current position is simply a matter of adding 3000 to the address pointer. However, it does have the problem in that the memory bus must be able to cope with resolving every possible address - with a 1 MB memory space, that requires 20 address bits (2^{20} = 1048576 bytes, or 1 MB). This has repercussions for the design of the computer's motherboard in that everything to do with the memory bus needs to support all the address bits at once.

By contrast, the 8086 performed things differently. Rather than imagining the boxes in a single row, rearrange them mentally so that they are now in a two-dimensional grid 64 KB wide. Each byte's address now consists of two parts, effectively the X and Y coordinates of the grid. The X coordinate, called the segment address by Intel, selects which 64 KB row to deal with. The 8086 had a 1 MB memory space, so there were 16 segments. The Y coordinate - the offset in Intel parlance - specifies which byte in that 64 KB segment to address.

This has two effects - the segment can be stored in 4 bits and the offset in 16. Although this adds up to the same 20 bits as the Motorola design, by having the address in two parts it is possible to deal with them separately. This means that the maximum address size that the memory hardware needs to deal with at once is only 16 bits. 16-bit support chips were already available and relatively cheap, as they had been developed for use with the 8080. There was also the bonus in that

"Intel realised that the 8080, though successful, wasn't good enough to compete with the new 16-bit processor (the 6800) being developed by Motorola."

the 8086 was largely machine-code compatible with the 8080. The plan was to allow multiple instances of CP/M to run on one computer, each in its own 64 KB segment.

Segments

There is one big disadvantage with a segmented memory architecture, though. Developing applications is made more difficult as the programmer needs to be constantly aware of the possibility of filling one segment with data and so having to alter both the segment and offset addresses to get to another one. The 64 KB segment size of the Intel processor has become deeply ingrained in the designs of the PC, DOS and even, to a certain extent, Windows. For example, the PC's DMA (Direct Memory Access) hardware, which allows expansion cards to directly access the machine's main memory without having to go via the processor, can only deal with 64 KB of data at a time. DOS's memory space is 10 segments in size - 640 KB. In Windows, DDE transfers (a method of sending data from one application to another) are limited to 64 KB chunks.

The 8086 and 8088 were, to all intents and purposes, internally identical. They ran the same code in exactly the same way. However, the difference between the two processors was that the 8086 had 16-bit external buses for data and addressing, while the 8088 had 8-bit buses. The trade-off was that because the internal processor was 16-bit, it took two cycles to get data into the processor with the 8088 (ie, 8 bits were transferred each cycle) whereas the 8086 could take a 16-bit chunk of data in one cycle. This resulted in the 8086 running approximately 150% to 200% faster than the 8088, although IBM used the

Pipelining

Modern processors decode and process each instruction in stages, each stage being dealt with by a separate section of the processor. In the i486 and Pentium, these stages are:

- Prefetch, where the instruction is brought from the cache.
- Instruction decode, where the type of instruction is determined, such as transfer of data from one place to another, a mathematical operation, a conditional jump and so on.
- Address generation. If the instruction uses a data address which is relative, such as "current address plus 50 bytes", the absolute address is calculated here.
- Execute. The instruction itself is executed.
- Write-back, where the results of the instruction, such as the sum of two numbers in an add instruction, are written to memory.

A conventional processor performs each of the stages one by one, and has to complete all the stages for one instruction before it can start processing the next one. Some of the stages aren't relevant to all instructions, of course - an instruction that takes two numbers already in the processor's registers, adds them together and puts the result in another register has no need for the address generation step. However, even if an instruction could be processed in less than one clock cycle (although most take several cycles), the prefetch stage only works each time the clock ticks. Therefore, the processor can only hope to complete one instruction per clock tick.

With the i486, things are performed slightly differently. As soon as one instruction has passed through prefetch, the next instruction is read in. In this way, instead of having large parts of the processor idle much of the time, it's constantly busy processing several instructions at once. This is known as pipelining. If one instruction doesn't need a particular stage, it skips it and all the subsequent instructions are shuffled up to take advantage of the idle processor section. This can mean that some instructions are processed in less than one clock cycle. Note that the pipeline only relates to the main integer processor sections; if the instruction is for a floating-point operation, it is turned over in its entirety to the floating-point co-processor.

Pipelining isn't 100% efficient, as sometimes the result of one instruction invalidates the next one in the pipeline - for example, consider a conditional jump instruction. If the result of the condition is true, then the next instruction to process isn't the next one in the pipeline, but whatever instruction is at the address that the jump moves to. Another example is where one instruction uses the result of the previous one as an operand. If, say, the first instruction adds two numbers together, then the next instruction takes this result and copies it to another part of memory, then the add instruction would have written the result to memory *after* the second instruction has gone through prefetch, and so is working with out-of-date data. In situations such as these examples, the pipeline is flushed - the results of any prefetch, instruction decode, address generation and such like are thrown away and the processor starts from scratch again.

The Pentium takes this even further. Instead of having one pipeline, there are two - one for each processor unit. These work the same as the i486's pipelines, in that as soon as one instruction has moved out of prefetch, the next one is read in. Of course, as there are two of them in parallel, the results of one of the processor units could invalidate the instruction that the other one is working on. As with the i486, in this circumstance both pipelines are flushed.

An efficient pipelined processor is, to a certain extent, dependant on the compiler making sure that the pipeline needs to be flushed on as few occasions as possible. Therefore, a compiler optimised for an 80386 processor needs to be altered slightly to get the best out of the i486, and altered again to get the most from the Pentium.

PC Processors

"The design of most PCs being produced today is such that the processor can be easily replaced to allow the machine to be upgraded."

latter in the first PCs as they could use cheaper support chips and were easier to deal with.

To complement the 8086 was an optional maths co-processor, the 8087. The standard processor could only cope with integer maths, so if you wanted to deal with floating-point numbers you either had to write the floating point routines yourself, or stick in an 8087 that could do them in hardware. Because the maths co-processor was always a very much optional device, the vast majority of applications are written to do floating-point maths in software as then they would run on any machine.

80286

The 80286 was quite an advance on the 8086/88. The memory address bus was expanded to 24 bits, giving a maximum memory capacity of 16 MB, although it was still a segmented architecture. It ran in one of two modes: when first powered up, it was effectively just a faster 8086, including the 1 MB RAM limit - this was called real mode. However, by switching into protected mode, the processor could access the full 16 MB of memory. Moreover, the segment boundaries could be moved - the maximum segment size was still 64 KB, but it did at least offer some flexibility.

The name "protected mode" came about because the segments could be set up to contain executable code, data or a mix of both. The reason for this is that any attempts to execute the data in a data segment as code would generate a protection fault, as would attempts to rewrite the code in a code segment. In this way, the processor could protect the memory against rogue programs that might crash a non-protected system - protection faults would be trapped by the operating system, and any application causing a protection fault could be shut down by the OS. This is vital for a multi-tasking system, so that one program cannot accidentally corrupt another one. In practice, however, the protected mode on the 80286 was rather clumsy so few operating systems used it to its full advantage.

The other serious disadvantage with protected mode was that, although there was a processor instruction to go from real mode to protected mode, there was no command to go from protected mode back to real mode. In some respects this made sense, as Intel expected systems that used protected mode to enter it and stay there. However, this created a problem on PCs in that the PC BIOS, the low-level commands stored in ROM that were used for interfacing with the hardware, couldn't run in protected mode, and neither could DOS itself. Therefore, to multi-task DOS applications, you had to switch back to real mode. This was a problem as there wasn't an instruction in the 80286 instruction set to go from protected mode to real mode.

The designers of OS/2 discovered that the only way to do this was to either reset or crash the processor, so causing it to come back up in its default real mode. Neither was particularly satisfactory or quick, but it did at least work.

The 80286 came in two IBM machines, the XT286 and the PC/AT. As with the 8086/88, the 80286 was an integer-only processor so if you wanted to perform floating point calculations in hardware, you needed a maths co-processor called the 80287.

80386DX, 80386SX And SL

Motorola, Intel's biggest competitor in the microprocessor market, had already come out with a 32-bit processor so Intel had to catch up. Also, the 80286 was widely regarded as rather flawed, so whatever Intel produced had to be good enough to get their reputation back. And so was born the 80386, later renamed the 80386DX.

IBM's 486 Processors

IBM has an agreement with Intel to allow it to produce compatible versions of Intel processors. The result of this is the 486SLC, and the 486SLC3, code-named "Blue Lightning". The 486SLC is a 3.3 volt 486SX processor with 386SL-type power saving facilities. Intel also produced a 486SL processor, but not for very long so the IBM version has appeared in many laptop and notebook PCs. The 486SLC has a smaller cache (4 KB) than the straight Intel 486SX processor so its performance suffers slightly.

Blue Lightning is more interesting. Again based on the 486SX, but this time with the full 8 KB cache, the 486SLC is a 25/75 MHz clock-tripled device that runs at 3.3 volts. As with the 486SLC, the SLC3 incorporates static electronics so the clock speed can be turned down to 0 MHz, and it can shut parts of itself down to conserve power. More important, for many people is that the Blue Lightning is available with a 16-bit external bus, retaining the 32-bit bus internally (like the Intel 80386SX). Moreover, it is a relatively simple operation to replace a 386SX processor with the SLC3 - Kingston Memory Inc is producing a device holding an SLC3 and power conditioning devices that clips on top of an existing 386SX and takes over control of the machine. Blue Lightning isn't as fast as Intel's i486DX4, but the upgrade possibilities are making it an interesting product.

"The standard processor could only cope with integer maths, so if you wanted to deal with floating-point numbers you either had to write the floating point routines yourself, or stick in an 8087 that could do them in hardware."

This had a 32-bit memory bus, giving a total physical memory capacity of 4 GB, although it could cope with 16 TB of virtual memory. Internally it was 32-bit throughout, and had a 32-bit external data bus. As with the 80286, when first powered up the 80386 ran in real mode, and was effectively just a vastly faster 8086 - 1 MB limit, segmented memory and all. It could also be switched into 80286-like protected mode, although segment sizes were now much more flexible and could be greater than 64 KB.

The 386's memory space could be set up so that it contained multiple 8086 memory spaces, and the processor could multi-task between them - this was virtual 8086 mode. Each virtual 8086 thought that its RAM ran from address 0 to 1024 KB, even though this 1 MB chunk could be anywhere in the 80386's memory space and could even be split up into several sections.

Finally, the entire memory space could be made flat - no segments at all. These memory addressing tricks were taken care of by a special memory management unit in the processor, that sliced up the physical memory as required. The 80386 also came with an optional new support chip that could control a RAM cache, so boosting performance even further. Finally, instructions were available to switch the processor from one mode to another as required, and much faster than the 80286 bodge of crashing the CPU.

In some respects, the memory management capabilities of the 80386 were more important than the fact that it was actually a 32-bit processor. For example, Windows 3.1 will run on a 16-bit 286 processor, and the extra facilities available when in 386-Enhanced mode are down to memory management. DOS applications are multi-tasked using the virtual 8086 mode, for example. The vast majority of applications in use today are 16-bit. It is only with the interest in OS/2 2.x and Windows NT that 32-bit applications are starting to become common for PCs.

Support Chips

Because of the 32-bit memory and data buses, the external support chips for a 386 system also had to be able to cope with 32 bits of data at a time; such chips had to be specially designed for use with the 80386. At the time, few people needed the power that the 80386DX processor could provide, so Intel did a similar job as it did with the 8088 - the 80386SX had 32-bit internals, but the external data and memory buses were only 16 bits wide.

The maximum memory capacity was also reduced to 16 MB. The 80386SX was a hugely popular processor, as it provided all the memory management facilities of the 80386DX and so could run all its software, but could use support chips that had already been designed for the 80286 - this allowed 386SX machines to be much cheaper than their DX counterparts. Of course, they were slower, but for many people this speed differential was less important than the price difference.

Maths co-processors are available for both the 80386DX and SX, called (predictably) the 80387DX and 80387SX.

The SL Range

A later development of the 80386SX was the 80386SL. Until the late 1980s, laptop computers were hampered somewhat by the quality of the screens available. However, with the advent of new technologies such as TFT (Thin Film Transistor) LCD panels, and even colour displays, the demand for portable PCs boomed. This was helped by the advances in battery and hard disk technologies, with the storage capacity in a given size increasing for both. This left the system's processor at something of a disadvantage - the standard 80386SX processor consumed enough power to seriously affect battery life. The 80386SL got round this by using static junctions, a type of electronics that can be slowed down to a standstill without losing its current state - standard dynamic electronics needs to be periodically refreshed or it forgets what it was doing. The 80386SL could also shut down parts of itself and the system as a whole when they weren't being used to further save power.

"Because of the 32-bit memory and data buses, the external support chips for a 386 system also had to be able to cope with 32 bits of data at a time; such chips had to be specially designed for use with the 80386."

PC Processors

i486DX

Most people were generally happy with the core facilities of the 386-range of processors, and a maximum RAM capacity of 4 GB would keep most people happy for a while. Therefore, when the replacement for the 386 was being designed, Intel kept the basic facilities the same and just made it faster. The i486 contained an optimised version of the 386's instruction set that had been rearranged to support pipelining (see the panel). It also had an, again, optimised version of the 80387DX maths co-processor on the same chip, and to top it all off there was also 8 KB of cache memory along with a cache controller. All this technology resulted in a chip with some 1.2 million transistors on it, and one that ran programs 150% to 200% faster than a 386 processor with the same clock speed.

i486DX processors are available that run at 25, 33 and 50 MHz. Despite the i486DX being much faster than the 386, fundamentally there's not that much difference between them - they run the same code, they can do the same things, it's just that the i486 does them in a shorter space of time.

i486SX

The relationship between the i486DX and the i486SX is quite simple - the i486SX is an i486DX without the maths co-processor. It has the same size address and data buses as the i486DX, ie, 32 bits each. If you want a maths co-processor in your i486SX machine, you purchase an i487SX and plug it in next to the main processor. This is, essentially, a standard i486DX processor, complete with maths co-processor, with a pair of pins swapped round and a facility whereby it turns off any i486SX processor that's in the same machine. Once you plug the i487SX in, it completely disables the i486SX and takes over operation of the machine for itself. As it's faintly ridiculous to have two processors in a machine, only one of which is doing anything, most manufacturers these days construct their i486SX machines such that you can take the i486SX out and replace it in its entirety with a standard i486DX.

"The relationship between the i486DX and the i486SX is quite simple - the i486SX is an i486DX without the maths co-processor. It has the same size address and data buses as the i486DX, ie, 32 bits each."

i486DX2

The i486DX2 was the first of the "clock-doubled" processors. The processor's internals ran at twice the speed of the external bus - a 33/66 MHz i486DX2 ran at 66 MHz internally, but at only 33 MHz at interface between the processor and the rest of the machine. The reason for this is that designing motherboards that run at anything above 30-40 MHz is neither cheap nor easy. At these sort of speeds, the designer needs to be very careful to ensure that adjacent data lines don't interfere with each other. High speed support chips aren't exactly inexpensive, either.

The drawback of clock-doubling is that there is a bottleneck between the motherboard and the CPU - although the CPU needs data at 66 MHz, the motherboard can only supply it at half that speed. The cache RAM on the processor helps to alleviate this to some extent, but a 33/66 MHz i486DX2 gives roughly the same performance as a 50 MHz i486DX. However, as it can provide this performance at a good deal cheaper price, few people are that bothered. In addition to the 33/66 MHz i486DX2, there is also a 25/50 MHz version.

Upgrades

There is another benefit of the DX2 - you can use it to upgrade existing i486DX machines. If you have a PC with a 33 MHz 486DX processor, then simply take that out, put in a DX2 and you'll get a 50% - 80% performance increase. To do this, the motherboard must support the new processor. These days, the processor is usually plugged into a ZIF (Zero Insertion Force) socket on the motherboard to make replacement easy. Intel has also taken to labelling the sockets so you know what can be plugged into it - a type 1 or type 2 socket can take an i486DX or i486DX2 processor.

i486DX4

Intel has recently announced the i486DX4 (normally just referred to as the DX4) processors. As with the DX2, the DX4 runs at a higher speed internally than it does at the external interface. Plug a DX4 into a motherboard running at 33 MHz, and the processor runs at 99 MHz internally (ie, it triples the external clock speed). As the mismatch between the external and internal buses is quite extreme, the internal cache has been doubled in size to 16 KB to help compensate. Interestingly, if you plug a DX4 into a motherboard running at 50 MHz, it only doubles the external clock rate to give an internal speed of 100 MHz. There is also a version of the DX4 for 25 MHz motherboards, that triples the internal speed to 75 MHz.

Power Levels

Most processors that Intel has produced run at five volts, as this is the standard voltage for TTL (transistor-transistor logic) chips. However, for the higher speed processors such as the DX4, five volts is a bit of a problem - to make the processor run faster, you reduce the size of the chip. The smaller

the chip, the quicker the data signals can get around. However, with a 5V supply, the heat that each electronic junction creates is so high that the chip can't easily get rid of it quickly enough.

Therefore, for the DX4, the power supply voltage has been reduced to 3.3 volts. This reduces the power consumption (and hence heat generation) enough to allow the smaller chips to function without damage.

Unfortunately, the 3.3 volt supply creates a bit of a problem for motherboards that, up until now, had expected their processors to run at five volts. This is eased somewhat in that the DX4 can cope with support chips supplying data at 5V, it only needs the 3.3V at the main power supply pin. Therefore, motherboards are being produced that can supply either 5V or 3.3V to the processor at the flick of a switch (these are machines with a type 3 processor socket), alternatively some third-party companies will be producing daughterboards that plug into type 2 sockets and carry a DX4 along with a power regulator to cut down the 5V supplied by the motherboard to 3.3V for the processor.

The DX4 processors also incorporate the static electronics and power shut-down facilities of the 386SL, and this allied with the lower power requirements of 3.3V chips means that it's only a matter of time before DX4 processors start appearing in laptop computers.

Pentium

The Pentium is the latest in the line of Intel 80x86-compatible processors. It's a semi-64 bit device, with 64-bit wide data buses both externally and internally, feeding 16 KB (8 KB for instructions, 8 KB for data) of 64-bit cache memory which in turn feeds twin integer CPU pipelines. Each CPU pipeline is effectively a slightly faster version of the 486's pipeline, so although the Pentium can move 64 bits of data around at a time, the actual processors are 32-bit devices that operate in parallel rather than true 64-bit units such as the Digital Alpha processor. Also hanging off the main data bus is a single, optimised version of the 486DX's floating point co-processor unit.

As described in the box on pipelining, running twin processor units is not totally efficient as there are occasions where the result of the instruction being processed in one unit invalidates that in the other, causing the pipelines to be flushed. In practice, this means that a Pentium runs around 50% faster than a 486DX with an equivalent clock speed. However, this is dependant on compilers making the most of the Pentium's facilities, so it is possible to get higher speed differentials on some operations.

The first versions of the Pentium came in 60 and 66 MHz versions, had 3.1 million transistors and ran at 5V. The heat generated by these devices is considerable - the 66 MHz version can give out 16 watts. This created a cooling problem in many PCs, so most Pentiums are now supplied with a heatsink and small fan mounted on top. Similar devices are also being used for DX2 and DX4 486s, although in most cases just a heatsink on its own is sufficient for these.

Intel has recently announced the next generation of the Pentium. The 90 MHz Pentium runs at 90 MHz internally and 60 MHz at the external bus, while the 100 MHz Pentium runs at 100 MHz internally and 66 MHz externally. Although the new Pentiums run at the same external speed as the old ones, you cannot just take out an old one and put in a new one. First, the new ones run at 3.3V (and incorporate SL-type power management facilities). Second, the new ones have additional functions that require more pins. These functions are based around multi-processor facilities - there is now arbitration logic to allow two Pentiums to access the same second-level cache memory without conflict.

Additionally, there is a multi-processor interrupt controller built onto the processor's silicon that can support up to 60 Pentiums connected together symmetrically. (IBM has already announced that it's working on an SMP (symmetric multi-processing) version of OS/2. Microsoft is rumoured to be doing the same with Windows NT.)

Conclusion

Intel's 80x86 range of devices has been the most successful family of microprocessors ever. Quite where Intel can go from here, though, is open to question. The major benefit of the 80286 and onwards has been the compatibility with 8086 software; these days, however, this is becoming something of a hindrance. That being said, the power management and multi-processor facilities of Intel's latest processors shows where the company thinks things are going - "green" computing, where the amount of power that even desktop PCs consume is reduced, and multi-processor computers for operating systems such as Windows NT.

PCSA

The Author

Peter Matthews is a writer and journalist specialising in PCs and networking. He has previously worked as PC Support Manager for a major international company.

Tuning Windows For Winword 6

Microsoft Word for Windows 6 is bigger and slower than Version 2. Roger Dalton presents some tips to speed up the program's execution speed.

Version 6 of Word for Windows is probably the most complex word processing application available on the PC platform. It's also, unfortunately, one of the largest. A full installation occupies some 20 MB of hard disk space, though you can reduce this to around eight if you manage without the help files or any of the advanced features.

Microsoft has recently introduced a set of patches for Winword 6.0, taking it to 6.0a. If you have single-user versions of the program, you can download the patch (some 700 KB) free of charge from the Microsoft areas on CompuServe, or get them free from your local Microsoft office. There are separate versions of the patches for the UK, USA and Australian versions, so make sure that you download and install the correct one. If you don't you'll end up with an unusable installation and you'll have to re-install Word from your master disks.

If you have a network version, you can't download the patch file. You have to contact Microsoft for a complete set of 6.0a disks. These are free of charge, too.

This article offers some hints and tips for speeding up Word for Windows 6.0 and 6.0a.

Figure 1 lists the major problems that are fixed in 6.0a. If you don't suffer from any of these, it's not necessary to install the patch though it's a good idea to do so.

WINWORD6.INI Settings

To increase bitmap-redrawing speed and scrolling speed in Word, add the following two settings to the [Microsoft Word] section of your WINWORD6.INI file, located in the WINDOWS directory:

BitMapMemory= sets the amount of memory (in KB) reserved for cache memory for bitmaps. Increasing this number increases the size of the bitmap cache that Word uses for redrawing pictures quickly.

The BitmapMemory setting should not exceed the amount of available free RAM. A setting of 1024 should provide enhanced performance in

Disk Error While Saving, Embedded Object Becomes a Box
When you move an embedded object within a document using either cut-and-paste or drag-and-drop, and you then (full) save the document, the object may change appearance and will be uneditable. The object usually becomes an empty square box. Subsequent save attempts will result in "Disk is full", "too many edits", or "operation will be incomplete" error messages.

Problems Caused by SHARE.EXE
Word 6.0a includes an updated VSHARE.386 file. This revised version replaces SHARE.EXE. This version of VSHARE.386 provides greater compatibility with many applications running under Windows 3.1 or Windows for Workgroups 3.11 and later. To run Windows in standard mode, you still need SHARE.EXE. To run in other modes, it's no longer necessary to load SHARE.EXE.

"General Mail Failure" Error Editing Routing Slip on a Mail Gateway
If you edit a routing slip in a Word document, a "General Mail Failure" error message may occur.

No kerning with Adobe Type Manager Fonts
ATM fonts do not print kerned, because ATM fonts use an older application program interface (API) for kerning; this API is not supported in Word 6.0. Word 6.0a supports the older API.

"Not available" Error Messages with Adobe Type Manager Fonts
Word displays the following error message for a group of Adobe Type 1 fonts, even though they are available on the system: The font ' <font name>' is not available on your system. Do you want to use it anyway?

No Anti-Aliasing for TrueType Fonts
On a computer running an ATI Crystal Font video driver (or similar driver), the Crystal fonts are displayed with clean, nonjagged edges due to anti-aliasing. TrueType fonts, however, are displayed with jagged edges because no anti-aliasing occurs.

Can't Remove Space Around Footnote Reference Marks
By design, Word 6.0 places white space after automatic footnote reference marks. In response to customer requests, Word 6.0a inserts the white space as a regular space character to the right of the footnote reference mark, which means you can edit or remove it if you want.

Dr. Watson Errors in Examples and Demos
The Dr. Watson program displays errors when you choose the EXAMPLES AND DEMOS command from the Help menu.

Cannot Remove Borders on Pictures
Sometimes you cannot remove a portion of a border from an inserted picture or object, even when you choose None in the Format Borders And Shading dialog box.

Figure 1 - Major Fixes in Word for Windows 6.0a

Word (256 is the default setting). You can insert the command anywhere in the [Microsoft Word] section of your WINWORD6.-INI file using the following syntax:

[Microsoft Word]
BitMapMemory=xxxx

CacheSize= sets the amount of memory (in KB) reserved for cache memory for Word documents. The default value for the CacheSize command is 64, or 64 KB. Increasing this setting (in multiples of 64 KB) improves the speed of scrolling, searching and replacing, the GO TO command, and document load and save times. If your system has plenty of memory and you work with many large documents, consider setting the CacheSize to 256 KB or 512 KB.

Cross-References Updated Incorrectly
Bookmarks for captions change (shrink) to include only the caption label (instead of the full text of the caption) when you insert new captions later in the document. When you update cross-references to these captions, the cross-reference includes only the caption label.

SEQ Field Displayed Incorrectly
The result of a SEQ (sequence) field in the header and footer pane displays "1" or "0" instead of the correct number in the sequence. Although the displayed version of this field is incorrect, the printed versiof is correct.

Password Displays in Converted File
If you save a password-protected document in another file format, the password appears as text in the converted file.

GP Fault Creating New Document Based on Template With Link to Excel
When you start Word with the /n switch, a general protection fault occurs if you create a new Word document based on a template that contains a link to a Microsoft Excel spreadsheet.

Fonts Printed Without Kerning
Fonts do not print kerned in point sizes larger than the Kerning For Fonts setting in the Format Fonts dialog box. Fonts print kerned only if they are formatted with the point size specified by the Kerning For Fonts setting.

Word 2.x Mail Merge Document Loses Data Delimiters in Header
When you open a Word 2.x for Windows mail merge main document, and if the data document is not a Word 2.x table, Word 6.0 removes the delimiters from the header record, which concatenates all the field names in your data document. In other words, Word reads the header record as a single field name (for example, NameAddressCityZip).

Different Hierarchy for Choosing Which Same-Name Macro to Run
Word 6.0 follows a different set of rules than Word 2.x to determine which template macro to run when two macros have the same name: Word 2.x runs the macro from the template attached to the document in which the insertion point is currently located, while Word 6.0 runs the macro from the template from which you launched the macro.

Cannot Open Password-Protected Paradox 3.5 Files
A "File Access Denied" error occurs and you cannot open (via ODBC) a password-protected file created in Paradox version 3.5 or earlier because Word 6.0 does not prompt you for the password. You can open these files in Word 6.0a because the ODBC driver prompts you for the password.

Insertion Point Scrolls Off Screen with WordPerfect Help On
If the WordPerfect Help option is selected and you press ENTER several times, Word 6.0 scrolls the insertion point off the bottom of the screen instead of advancing the screen.

Cannot Create Non-File Links
In Word 6.0, you cannot create a link to server an application that does not use a filename to reference data. For example, Word 6.0 is incompatible with a database program that references a database by name and does not use the name of the file.

PC Hangs Launching Word as Server when WordPerfect Help is on
If the Navigation Keys for WordPerfect Users option is turned on and you launch Word from another OLE 2 program, the PC may hang when you press the ESC key. Sometimes it takes a while for the hang to occur.

Figure 1 - Major Fixes in Word for Windows 6.0a (Continued)

Hardware Optimization

Because Windows 3.1 uses extended memory, the more extended memory available, the better Windows 3.1 performs. Installing additional extended memory helps improve performance.

There is a trade-off between speed and hard disk space, depending on your system's hardware capabilities. In either case, make sure that as much free memory as possible is available.

Because hard disks are often formatted with the incorrect interleave at the factory, setting the optimal hard disk interleave for your system will help to improve performance. Software such as SpinRite, can be used to set the interleave and some utilities can correct the interleave without formatting your hard drive.

If your system has a memory expansion board and the board can be configured as either expanded or extended memory, configure all of this memory as extended memory. You can then use EMM386.EXE to emulate expanded memory only as needed by non-Windows applications that require expanded memory to run. In this case, make sure that the expanded memory driver is placed before the lines that load HIMEM and EMM386.

Configuring MS-DOS

Microsoft recommends that you upgrade to MS-DOS 6.2 so that you can load MS-DOS into the high memory area by setting DOS=HIGH or DOS=HIGH,UMB in your CONFIG.SYS file. In addition, MS-DOS 6.2 allows for proper disk maintenance with the ScanDisk and DeFrag utilities included with MS-DOS 6.2.

Make sure the most recent version of HIMEM.SYS, EMM386.EXE, RAMDRIVE.SYS, and SMARTDRV.EXE are stored in directories specified in the command lines of the CONFIG.SYS and AUTOEXEC.BAT files.

Ensure that the command that

Winword 6

loads HIMEM.SYS comes before any commands that load any other applications or drivers that use extended memory.

The SMARTDrive disk-caching driver produces one of the biggest Windows 3.1 performance improvements. For this reason alone, load SMARTDRV in the AUTOEXEC.BAT file and allocate the largest amount of memory as possible. Ensure that you load SMARTDRV after any CD-ROM drivers, or the CD-ROM drive won't get cached.

Ensure that the InitCacheSize and WinCacheSize parameters are properly set in accordance with the amount of memory installed on the system. If SMARTDRV is being loaded with no parameters, in some cases (MS-DOS 6.2 and 6 MB or more of RAM), the default settings can use up to 2 MB for a cache size in MS-DOS and 2 MB in Windows.

Set FILES=60 in CONFIG.SYS file unless you have a software application that requires a higher setting. Set BUFFERS=10 in CONFIG.-SYS if you use SMARTDRV.EXE. Using a high number of buffers with SMARTDRV will decrease efficiency. If you are not using SMARTDRV, set BUFFERS=30. More buffers may improve disk access times but will use more conventional memory.

Load EMM386.EXE if you are running non-Windows applications that require expanded memory or if you want to allocate UMBs for loading memory resident TSRs and drivers.

Load only the necessary TSRs, drivers, or software programs. Wherever possible, load TSRs and device drivers into the UMBs.

If the environment space is set by a SHELL statement in the CONFIG.SYS file, you can specify a smaller environment. Remove any commands for mouse drivers in your AUTOEXEC.BAT and CONFIG.SYS files if you only use the mouse in Windows and don't want mouse support while running non-Windows applications in 386-Enhanced mode.

Memory optimization plays a major role in system performance. If only Windows-based applications are run, don't load any drivers, programs, TSRs, or MS-DOS itself into the high memory area. Loading items into the high memory area frees up only conventional memory and if MS-DOS-based applications are not run, then conventional memory can be used to load these items.

Hard Disk Performance

Delete unnecessary application and system files, including .BAK and .TMP files. The fewer files your hard disk has to sort through, the quicker the access time. Make sure Windows is not running when you delete .TMP files.

Run ScanDisk or CHKDSK frequently to find lost clusters. If lost chains or clusters are detected, fix them with ScanDisk or run CHKDSK /F.

Use a utility program such as MS-DOS DeFrag to compact your hard disk regularly. A fragmented hard disk greatly impacts the performance of Windows, especially if SMARTDrive is installed or if you're using a temporary swap file. Make sure Windows is not running when you run the disk-compacting utility.

Configuring Windows

Use a colour or a pattern for the desktop background instead of wallpaper if you need to free memory for running applications. Bitmaps consume more memory.

Choose the lowest-resolution display driver that will meet your needs. In general, use the standard VGA driver to ensure faster display performance (but lower resolution and less colour support).

If you run in standard mode, set the application swap file to the fastest hard drive by setting the SwapDisk=

GPF Updating Large Number of Links
A GP fault may occur when you update a large number of links (typically from 80 to 200).

"Out of Memory" Error Opening Microsoft Excel 5.0 Chart
When you open an embedded Microsoft Excel 5.0 chart using the OPEN command (to edit the chart in a separate Microsoft Excel window), an "Out of memory" error message occurs. This does not happen when you edit the chart in place within Word.

Key Assignments not Saved in Templates
Word does not save template key assignments to the TAB and SHIFT+TAB keys.

GP Fault Running Find File Command
If you start Word using the /n switch and then open the Find File dialog box (by choosing Find File from the File menu), a GP fault occurs if the View option is set to Preview in the Find File dialog box.

Wrong Chapter Numbers on Last Page of Document
The chapter number always displays as "1" on the last page of your document in page layout view, regardless of the actual chapter number.

Wrong Chapter Number with Page Break Preceding Heading Paragraph
When your document includes chapter numbers in the page number, if a page break immediately precedes a heading paragraph that marks the beginning of a chapter, the wrong chapter number appears in the header or footer on that page.

Error Messages About Printer Port
By design, Word 6.0 confirms the validity of a port before printing. If there is no port present, Word displays an error message. With some configurations, however, Word incorrectly concludes that the printer port is not present and displays a "Local Port Not Present" error message.

GPF with INCLUDETEXT Field
If you insert an INCLUDETEXT field that references a bookmarked table cell in another document, a GP fault occurs. This error does not occur if the bookmark is not in a table cell.

DATE Field Result Clipped in Invoice Template
The result of the DATE field in the third column, first row, of documents based on the Invoice template (INVOICE.DOT) is truncated on the right side when you print it.

Figure 1 - Major Fixes in Word for Windows 6.0a (Continued)

entry in the [NonWindowsApp] section of the SYSTEM.INI file.

If you run in 386-Enhanced mode, create a permanent swap file on your fastest hard drive by choosing the 386-Enhanced icon in Control Panel and filling in the options in the Virtual Memory dialog box. The only reason not to use a permanent swap file is if hard disk space is at a premium. A permanent swap file, or virtual memory, generally allows the system to run faster because it uses contiguous disk space. A temporary swap file attempts to use contiguous disk space, but because of its dynamic nature, it can't always do so.

If you don't set a permanent swap file for 386-Enhanced mode, set your temporary swap file to the fastest hard drive by setting a value for PagingDrive= or PagingFile= in the [386Enh] section of the SYSTEM.INI file.

Windows For Workgroups

To optimize the performance of Windows for Workgroups on a workstation that is sharing resources, do not use a screen saver. Screen savers can degrade performance on a workstation that is sharing resources, so you should not use a screen saver, or you should use a non-CPU-intensive screen saver, such as the Windows Marquee screen saver.

If your workstation is only sharing resources (that is, it is only used as a dedicated file or print server), you can adjust the Performance Priority slider bar, using the Network icon in the Control Panel, to allocate more CPU time to the sharing of resources.

If you are sharing a locally connected printer, Print Manager (considered to be an application), must be running on the workstation. In this case, you must leave some CPU time allocated to applications so Print Manager can run.

If you are using a separator page for print jobs, use a less-complex separator for faster printing. To modify the separator page, open the Options menu in Print Manager.

32-Bit Disk Access

If there is an option to use 32-bit disk access in the Virtual Memory section of Control Panel, this means that your hard drive controller is, or appears to be, WD1003 compatible and can use this option.

32-bit disk access replaces the disk BIOS. It serves as a device driver that talks to the hard drive controller, watching for special calls (Int 13h) and handles them directly, bypassing the BIOS. In 386-Enhanced mode, 32-bit access speeds up disk access.

There are risks involved when using 32-bit disk access. There are some cases where this can crash the system, usually cases where the hard disk controller appears to be WD1003 compatible, but is not. Also, there can be some problems even on compatible controllers on some portable computers, specifically those that power down the hard disk to conserve power without telling the running software.

Permanent Swap File

Because of the calculation that Windows uses to set a swap file size, it is imperative that as much contiguous disk space and extended memory as possible is free and available. If a permanent swap file already exists, you may be able to increase the size.

Figure 2 lists the steps involved in doing this.

PCSA

1. In Windows, choose the Control Panel icon in the Program Manager (usually in the Main group).
2. In the Control Panel, choose the 386-Enhanced icon.
3. Choose the Virtual Memory button, and then choose the Change button.
4. Under New Swapfile Settings, in the Type box, select None.
5. Choose OK and choose Yes when asked if you are sure.
6. Choose Continue when asked whether to restart Windows or continue.
7. Choose OK in the Virtual Memory dialog box, then choose Exit from the Settings menu in Control Panel.
8. Open the Startup group and temporarily move any items to another group or rename the Startup group.
9. Quit Windows.
10. At the MS-DOS prompt, run a disk maintenance utility such as CHKDSK, or if you're using MS-DOS 6.2, ScanDisk. If errors are reported, convert the lost clusters or chains to files. After either of these utilities has run, you can check the .CHK files and delete those that are not needed.
11. Run a defragmenting utility such as Norton Speeddisk or PC-Tools Compress. If you are using MS-DOS 6.0 or higher, run DEFRAG. This will defragment your hard drive, creating more contiguous disk space.
12. Use a text editor such as MS-DOS Edit to modify the CONFIG.SYS and AUTOEXEC.BAT files to include only those device drivers and TSRs that are needed to run the system and Windows. Do not load MS-DOS or any other items into the high memory area. This will free as much extended memory as possible the next time you start your computer.
13. Using a text editor, modify your WIN.INI file (located in the Windows directory) and "remark" the LOAD= and RUN= lines by placing a semicolon at the beginning of each of those two lines.
14. Reboot the computer and start Windows.
15. Choose Control Panel from Program Manager.
16. Choose the 386-Enhanced icon and then choose the Change button.
17. Select a drive in the New Settings section that will net you the largest maximum recommended swap file (a permanent swap file cannot reside on a compressed or stacked drive).
18. If there is a check box at the bottom-left side of the dialog box labelled 32-Bit Disk Access, select that option.
19. Choose OK and choose Yes when asked if you are sure.
20. Choose Yes when prompted and informed about using 32-bit disk access.
21. When prompted to restart the computer or continue, choose Restart the Computer.

Figure 2 - Increasing the size of the Permanent Swap File

Acknowledgement

This article is based on official information supplied by Microsoft.

Lotus: Strategy For The Future

We present the first in an occasional series that examines the product development strategies being followed by major hardware and software companies. In this first article, Chris Long looks at how the Lotus software of the future will work together, and how it will integrate with the applications that are available today.

If there has ever been one company that has been on the computer industry roller coaster ride it is Lotus Development. Back in January 1983 when Lotus 1-2-3, the industry's favourite spreadsheet, shipped, the market was less than 18 months old. And when it hit the top of the best-seller lists in April of that year Lotus looked all set to take over the world.

The success of 1-2-3 was phenomenal. It was fast and simple to use, there was nothing else like it on the market, and it effectively assured Lotus's success for years to come.

This success continued to almost the end of the decade, though despite ending 1988 with revenues of $468 million, Lotus wasn't as healthy as it might have been. By June 1989, 1-2-3 version 3.0 was on the shelves, but it was over a year late, and when it hit the streets its reviews were disappointing. By now companies like Microsoft and Borland had been moving in on the spreadsheet market with Excel and Quattro. But when Windows 3.0 was launched in 1990 and went supernova, all Windows programs were dragged along with it and immediately Excel became the program to beat. This time the writing was on the wall for the DOS products.

Again Lotus struggled, this time to get a Windows spreadsheet through the door. 1-2-3 for Windows wasn't the success it could have been and again Lotus was slow. In 1991 it launched Improv for the NeXT computer which was a innovative new approach to the spreadsheet, but it wasn't to be available for the PC until 1993.

Notes

It was the launch of Lotus Notes that saw the company move back to a "power" role in the industry.

Notes is important to understand because it sums up the future of Lotus's role in the computer industry. And not only is Notes Lotus's future, it is strongly argued by people in the industry (enthusiastically applauded by Lotus) that it is the future of the industry itself.

Lotus Notes is essentially a database that works across a network that replicates itself (as parts or as a whole), which allows people to access any of this information (ie, a distributed multi-user database).

All this is built on a platform that allows it to serve its clients (users and other servers) in close to real time, across vast wide area networks using local area networks and telephone links. It runs on most high-end network operating systems: Novell NetWare, OS/2, Windows, Microsoft LAN Manager and Vines. Lotus is bullish about its cross-platform support, and refuses to promote any single platform as "better" than another, although it does admit that OS/2 2.1 does give "slightly better" support to Notes.

Lotus has also committed to support Microsoft Chicago (Windows 4) and is confident of having a module to support it soon after its launch. Lotus also supplies Notes client and server packs for Sun Systems as well as a Macintosh client pack.

"Notes is important to understand because it sums up the future of Lotus's role in the computer industry - not only is Notes Lotus's future, it is strongly argued by people in the industry that it is the future of the industry itself."

Public And Private

The database includes "published" information (anything that is publicly accessible) and private data (email etc). All the information needed locally is replicated across the various servers, and as individual users log on to the system, their version of the database (the local information) is updated from the server. Obviously this is a boon for people on the move

who can upload their own finished work completed "on the hoof" and download any new information (email messages, documents sent for approval etc) to their new location.

Thus Notes keeps all the different copies of the data, along with all the users on the system, synchronised by constantly updating its information. But it isn't simply connecting individuals. As technology progresses, so does the ability to have clusters of users working in groups. And this is where Lotus, with an established base in the Notes camp, will be going in the next few years.

Structures

Traditionally companies have worked on a hierarchical system of boss and subordinate, but more and more businesses work in "teams" where ideas are input and spread around the team, growing and diversifying in the process. This flat organisational structure is a group. There is no fixed number of people involved and, furthermore, this group may not even be in the same building or same country. Thus the technology needs to be able to support the topological diversity of its users - and quickly.

This is all the realm of groupware. Lotus has thought hard about this and it is hardly surprising that it will be using Notes as an integral part of how it intends to develop in the groupware arena.

Third Wave

At the moment Lotus argues that we are living in the third wave of computing. The first two were mainframe computing, where big computers automated clerical functions like accounts and general ledger functions. The second was personal computing that allowed executive management to participate directly with computers. It gave them access to, say, a spreadsheet or word processor, but all the same it still wasn't strategic computing, it was personal computing. All the actions were focused on the job at hand, and not in providing a company-wide solution.

The current wave is where, finally,

Lotus Notes release 3.0
The jewel in Lotus's future crown. This version was launched in August 1993 and the next version (Notes 3.1) is likely to be available 3rd quarter 1994. It will mainly address user-oriented enhancements. Lotus recognise it has to keep ahead with Notes thus it is attempting to fold as much user feedback in to the product as possible. Notes is also being positioned as a solution in more and more vertical markets along with specialist software.

Lotus cc:Mail release 2.0 for Windows
The current version was launched August 1993 and there will be a major upgrade by the end of 1994. It has just been optimised to run in the Windows 3.1 environment. Lotus intends to bring this product in line with Notes, which means making it a full client/server system. This means there will be both client and server editions of the product.

Lotus SmartSuite 2.1 for Windows
SmartSuite is the Lotus office "bundle". It has the latest version of AmiPro, 1-2-3, Freelance Graphics, Approach and Organiser. This will be upgraded (probably to release 2.2) to include Approach 3.0 when it is released.

Lotus SmartSuite for OS/2 1.0
Next version likely to be released 3rd quarter 1994. This version will run on OS/2 2.1 and has the OS/2 versions of AmiPro, 1-2-3, Freelance graphics and cc:Mail. Lotus has made a point of saying that it won't be upgrading the OS/2 product just because it has upgraded the Windows product. Both versions have separate development cycles.

Lotus 1-2-3 release 4 for Windows
Current version launched in June 1993 currently at version 4.01 (Lotus Notes enabled). Latest version of the top line spreadsheet. Next version (4.11) likely to be available 3rd quarter 1994.

Lotus 1-2-3 release 3.4 for DOS
Current version launched mid 1993 this is meant to be the DOS equivalent of the Windows version and runs on 286 PCs and above. New version (4.0) will be released iminently and is likely to match the functionality of the Windows version. Lotus have even developed a graphical front end for the DOS version that looks like Windows and includes smart icons.

Lotus 1-2-3 release 2.4 for DOS
Current version launched May 1993. This is the everyman version of 1-2-3 it will run on any PC (including 8088) and doesn't need expanded or extended memory. This is the end of the developmental line for this version as there are no plans for a next version, although Lotus will be issuing maintenance releases when necessary. Also Lotus have committed to continue making the 2.4 file format upwardly compatible to all the other versions of 1-2-3.

Lotus 1-2-3 release 2.0 for OS/2
This version launched in mid 93 next version scheduled for 3rd quarter 1994. Lotus are likely to give the next version full 32 bit support.

Lotus 1-2-3 release 1.1 for Macintosh
At the moment Lotus is "currently examining the viability of the Macintosh environment". This is corporate speak for "We will definitely communicate with it, but might well not write applications for it any more."

Figure 1 - The future of Lotus products

computers and communications technology combine to enable strategic - or enterprise-wide - solutions.

Lotus insists that this wave has come a bit early. It argues that this migration is premature because the user interface in client/server systems for transactional processing are identical to the types of systems that were built in the era of the mainframe and minicomputer and as such don't help the user.

Workflow (replacing the usual "paper chase" within companies with an electronic document system) is a key issue here. It's argued that we need to change the way these systems are developed to incorporate the traditional paper-based systems. This means changing the way a company might use an electronic system and, as a result, produce a new form of computing which Lotus calls group enabled office applications.

Lotus sees users going from the standard email system to a much richer environment of information sending and sharing. Perhaps the biggest area of interest will be the transition from electronic data interchange, which is really just sharing transactional information, to rich exchanges of information - including specifications and customer support information between different enterprises.

Extended Enterprise

Thus the notion of an extended enterprise that uses electronics to electronically bind customers, suppliers, business partners in ways never before possible is introduced. This means that products introduced to this area will have to take into account "scaleabilty" where they can grow and shrink to accommodate different styles of work and different volumes of people. They will also have to support multi-platform systems so users aren't tied to one computer system. And this includes addressing the need to be compatible with different systems outside any one enterprise.

OpenDoc

Together Apple and IBM have produced what is effectively a superset of OLE 2, called OpenDoc. OpenDoc has

Lotus

Lotus Improv 2.1 for Windows
A true "next generation" product, currently finding a niche market, given its astonishing flexibility in manipulating data, as a database viewing tool while it competes again its more popular sibling 1-2-3. The next version is likely to be rolled out 3rd quarter 1994.

Lotus Approach 2.1
Version 3.0 of the Lotus database will be launched early in 2nd quarter 1994. There aren't any dates available for the next version. 3.0 will be Lotus Notes enabled.

Lotus AmiPro release 3
This version 3.01 (notes enabled) was launched in 1993. It is not expected to be updated until 1st or 2nd quarter 1995. But at that time a major upgrade is promised which will probably include some sort of implementation of the groupware philosophy being developed for Notes.

Lotus Write release 2.0 for Windows
The little friend to AmiPro (this is the original Ami produced by Samna - who also developed AmiPro - and came with the company when Lotus bought it). Its development is still closely tied to AmiPro's and we will see a new version of Write for Windows with the new AmiPro, again in the 1st or 2nd quarter 1995 time frame.

Lotus Freelance Graphics 2.0 for Windows
The current version (2.01 - Notes enabled) launched end of 1993 wont be upgraded until 1st quarter 1995 earliest, although it's likely to be a major release.

Lotus Freelance Graphics release 2 for OS/2
Like its Windows sibling, this won't be upgraded until 1st quarter 1995.

Lotus Freelance Graphics 4.0 for DOS
This version launched in 1992 is probably the last version of the DOS product line. Although Lotus are committed to regular maintenance upgrades (although how regular is anyone's guess)

Lotus Symphony 3.0
According to Lotus "The product that wouldn't die". Launched in 1992 this integrated package is rather long in the tooth, it has a 1-2-3 spreadsheet system and a WYSIWYG front end for charts. Lotus have no plans for an upgrade. Although Lotus does offer a special SmartSuite upgrade pack to convert Symphony files and, more importantly, macros in to 1-2-3 files and macros.

Lotus DocOnline
A CD ROM containing all the documentation for all Lotus's Windows products. Updated whenever a new version of a program is released.

Lotus SmarText
Lotus's document construction tool for developing complicated electronic text systems.

Lotus Organiser 1.1 for Windows
This version released in mid 1993 is slated for a major upgrade to version 2.0 by the end of 1994.

Lotus Agenda
Still with us, but without any real hope of further development. Maybe the odd maintenance.

Figure 1 - The future of Lotus products (Continued)

"Lotus sees users going from the standard email system to a much richer environment of information sending and sharing. Perhaps the biggest area of interest will be the transition from electronic data interchange."

a broader sweep because, unlike OLE 2, it isn't tied to one operating system (Windows). Again, Notes will be supporting this too, allowing the user much more freedom of choice over what platform to move to.

Lotus says that in the future a system like Notes will effectively ignore any underlying operating system and this will allow companies to set up enterprise-to-enterprise communications, cutting out many inefficient paper transactions. And the only way paper can be taken out of paperwork is by enriching the system that handles it.

Lotus sees the computer applications as sophisticated document editors. If you have a spreadsheet that needs to be sent to a group it can be broadcast on the network via Notes, but rather than have an application that will work on the document the user will fire up the Notes document and work on the data without having to pull it into an application.

Thus the document is a container for the data and will access the applications to manipulate it. So we will see a move to document-centric applications. By embedding objects using OLE 2 and then using Notes to transport that, Lotus sees the usefulness of distributed computing coming in to its own.

Application

The application will be an editor for Notes information, meaning that you could create 1-2-3 templates that could be used as a Notes form, or AmiPro templates, Approach templates or indeed any other application. So instead of using a spreadsheet to do your own expenses, you can do your expenses on line on a expense management reporting system, which is just a Notes document with 1-2-3 embedded in it. The 1-2-3-ness will be hidden from view because the Notes environment becomes the carrier for its functionality across the Notes system. And this goes for word processing or any application, because the application engine is enabled across the system.

Script

Soon, applications will be bound into the Notes environment by the introduction of a new object-oriented inter-application programming language called Lotus Script. The first iteration will be available in Lotus Forms and Notes VIP, in the next few months. It will then evolve into all other Lotus products over the following 12 to 18 months. Lotus Script will be a substitute (although not a replacement) for native macros in 1-2-3, Notes and AmiPro.

ScreenCam

Lotus is also developing a new application called ScreenCam which is, in essence, a screen recorder. It allows you to point, click and manipulate data on screen and record all the action, including speech. Which means training and demonstrations will become easier and ScreenCam libraries can be set up as on line help systems. Perhaps it will be most useful as a group device, as "live" documents can be annotated with ScreenCam files so comments on a document from remote members of a team can be collected and relayed back to its author. Lotus is suggesting this will give rise to a new term: GIM (Group Information Managers).

The Future

New terms or otherwise, obviously Notes is the key to Lotus' future, but only time will tell if "group systems" will be the key to ours. While the Notes world is attractive and in a lot of ways logical, it isn't here yet. OLE 2 is a lot of hard work and it isn't as broad as OpenDoc. Unfortunately OpenDoc has less momentum than OLE 2 and almost certainly will take longer to take off.

But new products like Lotus Script and ScreenCam give users the opportunity to build on an already well developed applications base. Although whatever the future, right now Lotus is no longer the white knuckle ride it was.

PCSA

The Author

Chris Long is an ex-mainframe operator, computer service engineer and programmer, and now writes about computers for a variety of worldwide newspapers and specialist magazines.

Understanding OCR

Tim Phillips discusses the OCR market, and explains what the currently-available technology can and can't do.

Document Image Processing (DIP) and Optical Character Recognition (OCR), by which printed text is automatically converted to text files on a computer, has been an elusive goal for software developers.

The problem is that no matter how good the OCR package is, less than perfect is not good enough. Results for OCR packages are like those for gymnasts: 9.8 is a disappointment, 9.5 and you should have stayed in bed.

For an OCR package, 95% accuracy means one letter in 20 is wrong, which means one letter every four words. So, on a 400-word page there may be 100 words recognised incorrectly. 95% accuracy sounds great; 100 mistakes per page sounds terrible.

Improvements

Technological improvements to OCR engines are pushing the average number of mistakes lower each year, with dramatic improvements currently taking place. Allied with more sophisticated scanners, faster hardware and better indexing, OCR might have made the breakthrough to being considered an essential office technology.

There is certainly the market for it: instead of creating the paperless office, the personal computer has instead made it easier to print. The joke in the document management community is that the paperless office is about as likely as the paperless toilet. Meanwhile, large companies are being swamped with paper: invoices, reports, research - all of which has to be stored, often in warehouses acting as vast paper repositories.

Profile

The most high-profile use of OCR is as part of a DIP system. In this, paper documents are scanned, the image is recognised and made into a text file, and the text file is indexed. This index can be searched for a match, which finds the OCR text file, which can then access the original image if required.

It is important to recognise is that the OCR element is only an engine to convert an image to something that can usefully be stored in a database. It is the ultimate black box - images in, text out.

Other applications call for a more specific OCR implementation - for instance the systems in banks that recognise and sort cheque numbers. Errors cannot be made in these applications, but the machines that do the scanning only look for one set of numbers which are printed in the same typeface in the same position every time. OCR for recognising numbers, or specific symbols, is almost infallible and need not concern us. We are more interested in boosting the performance of our black boxes, both in terms of speed and accuracy.

In terms of manpower, it costs $25,000 to fill a four-drawer filing cabinet with paper, and $3,000 every year to maintain it. In this context, it is worth getting your OCR right.

Technology Background

To understand how to maximise your return from implementing OCR, you have to understand how an OCR package works. There are two distinct methods of OCRing your document. A package can use one or the other, and sometimes they use both. That's not forgetting the artificial intelligence-based methods of enhancing OCR either. The technology is in a state of transition currently - crossing platforms, adopting new techniques and enhancing performance. The American Institute of Information Management estimated that the major OCR products have increased their accuracy between 1992 and 1993 by between 25% and 50%

Polyfont technology was the first OCR technology, and is still going strong in products like Xerox Imaging systems' ScanWorX. Polyfont recognition is equivalent to a child playing with building blocks. Each block is pushed against a shape in a board, and when one fits, it falls through. A polyfont OCR package stores a large number of fonts, and flicks through them to match against the text object it has isolated.

"For an OCR package, 95% accuracy means one letter in 20 is wrong, which means one letter every four words. So, on a 400-word page there may be 100 words recognised incorrectly. 95% accuracy sounds great; 100 mistakes per page sounds terrible."

"In terms of manpower, it costs $25,000 to fill a four-drawer filing cabinet with paper, and $3,000 every year to maintain it. In this context, it is worth getting your OCR right."

The advantage of polyfont technology is that it is extremely accurate with fonts that it knows. It doesn't make many assumptions before it matches, so it is "dumb OCR" up to a point. This is graphically demonstrated when you try to OCR a document in a font it doesn't know, where the accuracy declines to almost zero.

Polyfont demands a large amount of memory and processing speed too, especially when comparing text against multiple fonts. If it is playing at home, it is tolerant and easy to manage, but it doesn't travel.

The original polyfont technologies were a natural extension of early OCR schemes which have existed for a couple of decades, for instance, the systems which read cheque numbers. Typewriters were designed with the ability to type fonts called OCR-A and OCR-B which could be recognised in software; it was a natural building job to extend the software by adding the ability to work with more fonts.

Another problem with polyfont technology is its scalability. If it is checking against twice as many fonts, it will take twice as long.

Omnifont

The alternative method of recognition is called omnifont technology, and addresses the shortcomings of polyfont technology by recognising the basic shapes of a letter rather than the exact font. So it is more likely to confuse an "e" with a "c" than a polyfont process, but it will recognise the "e" in more typefaces.

Omnifont OCR, as used by market-leading Calera's products, uses more processing power and less memory. It is also more easily scalable when recognising a large number of possible fonts. The problem is that it simply doesn't do the basic job as well as polyfont OCR.

A rough rule of thumb is that if polyfont methods achieve 97% letter accuracy on a document, omnifont technology could be stuck at 95%. This sounds like a minor difference, but it represents two-thirds more letter errors and consequently a huge number more word errors.

The Top End

The cavalry arrives in two forms. The first is neural network based context-sensitive systems, which can spruce up some poor text into readable form. The neural network tries a polyfont match, then if it fails to get a sufficiently good result reverts to an omnifont method. The result is the best of both worlds, and the learning capability of a neural network produces dramatically better results as the system is used.

This effectively enhances the results on clear text to close to 100%, and is the major source of accuracy im-

SUPPLIER	PACKAGE
Caere Corporation	Caere OCR 109
Calera Recognition Systems	Calera MM600 mm24su
Cognitive Technology Corp	Cognitive Cuneiform
CTA Inc	CTA TextPert DTK 1.2.9
ExperVision	ExperVision RTK
OCRON Inc	OCRON Recore
Recognita Corp	Recognita Plus DTK
Xerox Imaging Systems	XIS ScanWorX API

LETTER ACCURACY (%)	BEST	WORST
Caere OCR 109	99.88	88.58
Calera MM600 mm24su	99.92	92.99
Cognitive Cuneiform	99.26	85.28
CTA TextPert DTK 1.2.9	99.51	83.37
ExperVision RTK	99.90	93.14
OCRON Recore	99.67	81.49
Recognita Plus DTK	99.67	84.42
XIS ScanWorX API	99.83	93.23

Figure 1 - Leading OCR packages, and ISRI Results

Understanding OCR

provement in current releases. It has been successfully incorporated in ScanWorX.

The other add-on you may consider would be some form of Intelligent Character Recognition (ICR). Just as OCR sees what it sees and ignores the overall sense of a document, ICR ignores exact matches, and tries to apply rules of spelling and grammar to an OCRed text stream.

Anyone who has used a spell-checker and a grammar-checker knows that this can be a either a blunt weapon or a double-edged sword, depending on how you mix your metaphors.

If you are scanning the pages of a novel or a report, it will be a positive influence, and certainly worth the processing time. If by contrast you are scanning a form or any document containing serial numbers or tables, it will destroy any meaning that the OCR has salvaged.

ICR is to be used with caution. It is now supplied as an option with many OCR systems. Knowing when to use that option can make a positive difference to your results.

The other potentially disastrous problem with ICR is that it is language-sensitive. An OCR will do a credible job in recognising German letters, but an English-language ICR will make an awful mess with your scanned text.

The Business Overhead

When looking for an OCR solution, costing involves a relatively minor investment in software, a flexible investment in hardware, and a major investment in an operator's time.

Again it is instructive to pursue a few back-of-the-envelope calculations. A typical DOS PC running Calera OCR software embedded in a document image processing application will OCR a page of text in about 30 seconds. This can be shorter for clear, sparse pages but can also stretch to minutes when recognising new fonts or third- and fourth-generation documents.

You will find OCR solutions rated in characters per second, like dot matrix printers. The 30 seconds represents 75 cps on a 2500-character page - that's about 400 words.

There is no way to cut the scanning time using software-only methods. You have to throw hardware at your application to boost your processing rate, of which more later.

If you defer your OCR process and set the software to scan for 12 hours continuously each day, six days a week, it will scan 2 x 60 x 12 x 300, or just over 400,000 pages a year. This seems like a lot, but if you think in terms of magazines, it is equivalent to about 12 feet of magazines on a bookshelf.

This may be sufficient for some users, but for large archiving operations you will need to run dedicated networks of PCs, or upgrade your hardware.

1. Handwriting

The most common problem of all - how to OCR forms, such as those which have been filled in by hand, or how to OCR signatures.
The solution? Don't even think about it. The OCR package which has acceptable levels of recognition for form-based handwriting has not been developed. If you must use hand-filled forms, then use multiple choice booklets and go for mark recognition rather than character recognition, as the best handwriting recognition is around 60% - far too low to be reliable. Cursive handwriting is essentially unreadable to a commercial OCR package. Most forms are cheaper to enter by hand, where the number of uncheckable errors will be reduced.
Signatures are unreadable, but most documents will have a typed name under the signature. Make this an essential part of house style, or introduce initials-based references for all letters if they are to be searched by name.

2. Slope

Faxes and scanners produce a slight slope which confuses OCR. There's no easy solution here, apart from avoiding the slope. Less than 8% slope will still be readable - the resulting text file will break lines half way across the page, but can be successfully indexed and so the image can be recovered. More than 8% and the OCR will have trouble recognising individual letters.
The only way to avoid slope is by ensuring that your scanner or fax feeds as accurately as possible. For any OCR application, the scanner should be fed automatically, and a automatic feeder hopper is an essential part of your purchase list. If your documents are all the same size, this should ensure less than 8% error.

3. Fax

Faxes have increased error ratios - 80% is a good success rate with a document that scored 95% before faxing.
The solution is to originate and receive as much fax as possible electronically. This will reduce the detrimental effect of fax slope (see above) and will avoid rescanning documents. Electronic faxes do not fade either, whereas those on thermal paper do.
Unfortunately, there is an upper limit of less than 90% on OCR success with faxes. This is because faxes have a maximum resolution of 200 dpi, compared to 300 dpi and upwards for a laser printer and 1200 or 2100 dpi for commercial printing. Some OCR packages can apply a "fuzzing" effect on letters, which will smooth the effects of 200 dpi faxes, but specks due to line noise will always be a problem.

Figure 2 - Ten common OCR problems

UNIX

This is the major justification for using OCR on a UNIX platform. The scalability of UNIX means that it becomes possible to scale the hardware platform to fit your demand for OCR, and currently you can achieve dramatic benefits by using a workstation-based solution.

HP estimates that its 80 MHz PA RISC workstation, costing around $6000, would process one page every five seconds. The equivalent performance plus the cross-platform capabilities of PowerPC makes the platform potentially the price-performance winner, if you are using a standalone solution.

Alone among the OCR players, Calera boosts performance by offering hot boxes as dedicated OCR engines. While OCR's concept is that of a black box - you put in an image, it does something remarkable, and you get text out, Calera has taken the process further by supplying grey boxes, looking like mini-tower PCs. Conveniently labelled according to their speeds, the M/Professional card recognises 300 cps, the MM600 600 cps and the MM2400 recognises 2400 cps. All are SCSI II devices.

They work by processing using a dedicated processor, and by doing the job in parallel. In reality, the MM2400 is four MM600s working in parallel, but this is a problem that lends itself well to parallel processing.

With the MM2400 costing around $30,000, this is not for occasional users. It is wise to make a trade-off between processing power on your central CPU and processing power from your dedicated hot box. Currently this seems to reside at about the 600 cps level. If you want greater speeds, then it is cheaper to buy dedicated OCR hardware. If less than 600 cps is sufficient, then the cheapest way to get improvements up to that speed it to upgrade your CPU.

Before You Scan

There are several ways you can enhance your in-house documents to help an OCR read them effectively. Properly followed, these guidelines will have minimal effect on house style but a dramatic effect on the quality of your stored OCR text files.

First, keep your in-house fonts simple, and restrict your documents to as few fonts as possible. Retain spacing between letters and lines - this means resisting the temptation to track text too tightly or squashing lines of text to each other.

Small text is more likely to be misread too as a result of the printing process which can miss small loops or

4. Dot Matrix Printing

Dot matrix printers produce documents which are unreadable to an OCR engine. Again, the only solution is "fuzzing" the image to get round the unconnectedness of dot matrix dots. This will produce acceptable results with good-quality 24-pin printing, but 9-pin printing, as most line printing would be, is a virtual non-starter for OCR.

5. Pictures

There is no way to OCR a document so its picture information is intelligible. The solution is manual indexing and categorisation of pictures. This is the method used by all large picture research applications currently in operation. There may soon be a neural network based solution which can recognise pictures by pattern matching.
All OCRs can recognise pictures and will leave a gap in the ASCII text file where the picture should be.

6. Tables

The lines in a table confuse an OCR which interprets them as part of a word. Horizontal lines should cause no problem and will safely be ignored. Vertical lines passing too close to letters may spoil the OCR's effectiveness for an entire line - not good for table headings. In creating the original document, centre table text if possible to keep it away from the table lines. Also make sure that descendeurs do not touch the horizontal lines. Printing table lines in grey will improve performance too. Avoid shading in tables at all times.

7. Colour

Coloured text is ignored by OCR systems. If you wish to scan coloured text, photocopy the text or scan in black and white with no grey scales. OCRs do not register colour as they encode text as a series of whites and blacks to make binary strings for matching. Most coloured text could neither be white nor black in the eyes of the OCR - analysing coloured text would impose too much of a system overhead to be practical anyway. Dark colours will OCR as well as black text, provided the scanned image has sufficient contrast.

8. White On Black Text

White text reversed out of a black box will not OCR. If the text is a whole page, then scan a reversed image. Be warned that the spread of black ink on absorbent paper in the original will make the results of the subsequent OCR disappointing, as the characters will be more spindly than usual. If the text is a crosshead on a page or the title bar in a table, an OCR program will not read it; instead it looks to interpret the black box as a large oblong letter, and fails.

Figure 2 - Ten common OCR problems (Continued)

Understanding OCR

fill what should be holes with ink. Any letter touching its neighbour in any direction will also be unreadable.

Providing it follows these rules, italic text should contain no more potential OCR errors than Roman text, and bold text, provided it is not allowed to run together, has the advantage that it will scan as unbroken lines and so be more accurate.

For in-house printing, typing or faxing, try not to use absorbent paper, as this provides washed-out samples with ink which may seep into neighbouring characters. The type of cheap, absorbent paper used for plain-paper faxes may save material costs but is a definite source of OCR error.

Text which runs around illustrations is not usually a problem, although unusual layouts might not scan accurately, and the text file produced may have too many word breaks or discontinuations to index well. Coloured text may not scan with enough contrast to satisfy an OCR that it is looking for black on white text.

Tables should avoid close-running black lines and vertical or diagonal text.

This will also give you an impression of how well an existing archive will OCR, or the quality of text you will get from incoming documents. Where possible, avoid these errors by reducing the number of generations which a document goes through before OCRing. For example, a printed word processor file is a first generation document. Photocopying it makes it a second generation document, and faxing it adds another generation.

Photocopying the fax makes the document a fourth generation image, and your OCR would be lucky to achieve better than 80% from even the best original.

Avoidable errors can be reduced by not OCRing photocopies unless absolutely necessary, by printing on laser printers wherever possible, and by having non-urgent communications posted to you rather than faxed.

Finally, avoid unnecessary expense by substituting email or internet file transfer for paper document exchange whenever possible, and investing in electronic document management.

Avoiding OCR

It is possible to avoid OCR altogether in your DIP solution. If you are looking to classify your document by a series of fields which can be defined in advance, OCRing a document is just a way to give you every word you don't want as well as the ones you do. It will bloat your word index and possibly misread the key words you want to store.

AIIM estimates that it costs between $2 and $20 to index a document by letting a trained operator compile a keyword index in conventional database format. This is actually quite cheap for some repetitive documents - forms for instance. Unfortunately a keyword index is not sensitive to changes in emphasis over time.

If you index this article, ie the one you are reading now, you are unlikely to choose neural networks as one of the top 20 fields. Two years from now, neural networks will be vastly more significant than they are today but your index would not point towards their use in OCR. Not for this article, anyway. At least when you index a document manually you avoid OCR error.

Automatic indexing will index every word, including those with errors. This will usually not be important, as a reference word will crop up more than once in a document. A full text index of a recognised document can be as large as the text database itself. The implication for your storage is frightening, as you will have to have storage for your image, storage for your recognised text file and more storage for the index. Also, a large index takes time to search.

Currently the best indexing software is supplied by Excalibur technologies. Called Electronic File system (EFS), it indexes letter pairs throughout the document to provide an index which stores patterns of letters, and uses a neural network to provide "adaptive pattern matching". This produces an index which is tolerant of OCR error, as it will match "Tim Phillips" with "Tnn Pmllios" - a common enough OCR error pattern, but one which has destroyed my name for the purposes of conventional indexing. This type of approximate matching is extremely powerful and effective, and completely automatic.

Using it, the recognised test becomes a mid-point between the image and the index. You don't have to clean up an OCR file, so it is quick. It will also match sentences. The index is one-third the size of the text database.

9. Foreign Language Characters

When scanning foreign languages, the results are disappointing. Use an appropriate ICR package (see text) to interpret the results of the OCR match. This will improve the results. Make certain your package is looking for accented characters. On no account use an English ICR package with foreign text, as this will destroy what little sense remains. Other alphabets (Cyrillic, Hebrew, Greek) need their own OCR packages.

10. Line Breaks

OCRing a document does not produce continuous text, nor is it meant to, unless your document is a single-column, single-font piece of continuous prose. The OCR tries to keep the layout as close to the original design as it can. This is of course impossible, as you have just converted a typographical design to ASCII text in a constant size.
The retention of layout properties make correction and indexing easier to cross reference. Unfortunately it can mix captions in with text and make footers, page numbers and so on meaningless. The first set of considerations are commonly held to outweigh the second.

Figure 2 - Ten common OCR problems (Continued)

"Technological improvements to OCR engines are pushing the average number of mistakes lower each year, with dramatic improvements currently taking place."

The importance of fuzzy searching, as EFS's technique is called cannot be overestimated. Instead of breaking your back achieving half a percent improvement in your OCR accuracy, you can simply make your indexing tolerant of OCR error. Always think of your OCR package in terms of the process it helps - not as a product which has to be optimised at all costs.

The Legal Position

It's important to remember that OCRed text on its own has no legal status, and so you are never in a position to throw away the image of your document once it has been recognised. This is depressing on one hand: while your OCR text file will occupy perhaps 1200 characters per A4 page, this has to be linked to an image file, which even when compressed is many times the size.

On the other hand, it protects you as much as anyone else. Your text file can possibly be edited, and even if not may contain vital errors.

You may, however, be able to dispose of the paper copy of your image, but again, this is still a legal grey area. Provided the proper security is implemented on the image - that is, it is written to a WORM disc or some other non-editable medium, its text file will act as a useful pointer for your indexing program to search - and that is all.

In the UK, several large organisations are currently suing themselves over trivial lost documents in order to establish the precedent which will let them burn their document warehouses. As yet, the only major examples of paper documents being destroyed are in non-legal areas, for instance a reference library. Until the legal precedent is firmly established, beware of using OCR to link to a scanned image and then losing the link to the physical document as a result.

OCR APIs

Most of the time you will encounter your OCR software as an embedded package in an existing document management system. This is certainly not the only way to implement it though; you may wish to use an OCR package as a standalone, or else embed your own.

All the major packages are available as standalones, with a concerted effort to port to Windows over the last two years. Running under windows is a satisfactory solution, with the OCR package called up as necessary.

If your need is for a more streamlined approach, it may be necessary to tackle the programming work yourself. The good news is that the general standard of OCR APIs and SDKs is high, and there are few function calls to work with.

The APIs are generally easy to use, as an OCR does one basic thing many times. The only other calls you may need to implement are the controls, for instance turning on fuzziness when you need to OCR fax text. This gives a short list of function calls - for instance OCR_Doc (document.name) would be typical.

Major Products

There are eight major OCR packages commercially available. Most are DOS or Windows packages, but some suppliers - Caere and Xerox especially - have decided to concentrate on UNIX platforms, with the result that the package you buy for the PC might not have the bells and whistles you can get under UNIX. This has little to do with the operating system, and a lot to do with the release cycle. Nevertheless, all are available in some form on the PC.

OCR technology is commonly embedded, so fax packages for instance which contain OCR will not have reinvented the wheel. Instead, they bundle an optimised version of one of the packages listed in Figure 1. An example is the recently-released Delrina WinFax pro 4.0, using Xerox OCR routines to recognise faxes.

The most authoritative report on OCR accuracy is compiled yearly in the US by the Information Science Research Institute (ISRI), established with funds from the US Department of Energy. The annual research report is available by contacting the University of Nevada in Los Angeles. Be warned: it is not light reading.

The ISRI found in 1993 that for good quality documents, the overall difference in the scores of the best packages were not statistically significant when tested for letter accuracy. The results are best summarised in a table, and this is also shown in Figure 1.

The results show excellent results at best with a substantial drop-off for some packages for the worst documents. These sample results give only a flavour of the excellent results available in the full report, which deserves examination if you want a better picture of the options available.

PCSA

The Author

Tim Phillips is a freelance PC consultant and writer.

Troubleshooting Modem Sessions

If your users are having problems connecting to online services, this article should help them understand the problems and get everything working.

I can connect to a remote service OK, but once the connection is made I keep getting missing characters from within the text when I press Return to read the next line.

This is usually down to one of two things. First, and most likely, it's a parity mismatch. Parity checking looks at the integrity of received data to ensure it has been transferred correctly (although this is a very simple method and not to be confused with the error checking protocols supplied with most modern modems). Making sure you have the correct parity checking set is the answer to this problem.

The other cause of the missing characters could be a flow control problem. Software flow control uses the XON/XOFF setting, which sends a Ctrl-S to pause data and a Ctrl-Q to restart the data flow. This can sometimes cause problems when Ctrl-S characters are included in transferred data such as a binary file, which would make the transfer lock up due to the control characters being read and misinterpreted.

Hardware flow control uses the RTS/CTS setting and is accomplished by the RTS (Request To Send) and CTS (Clear To Send) pins of the serial cable sending the necessary pause and restart signals. Because of the limitations of software flow control when sending some binary files, RTS/CTS handshaking is always to be recommended. However, make sure that the cable between the computer and the modem has the RTS and CTS pins wired up or you will run into problems - a simple three-wire connection that only connects the transmit, receive and ground pins will not work correctly.

I can get connected OK, but then the screen just fills up with garbage.

This can be caused by a number of things. First check that the number of data bits and stop bits, together with parity checking, have been set properly for the remote system you are trying to connect to. Usually this will be 8N1 (8 data bits, no parity, and 1 stop bit), but some systems (including CompuServe) sometimes require them to be 7E1.

Second, check that the speed (measured in bps, though often incorrectly stated as baud) is correctly set for the remote system you are calling. Trying to connect to a modem that can only cope with 2400 bps when your modem is locked to 19,200 bps for example just won't work.

Enabling your modem software's "auto baud rate detect" feature will ensure that baud rates are agreed upon during the connecting handshake. Also ensuring that your modem is set up to negotiate the highest associated line speed supported by both modems. Using the Hayes command of "ATN1" should solve this particular problem.

The garbage could be a problem caused by line noise. Contacting your telephone service provider's engineering department for advice is the best bet in the case of persistent line noise problems. Ask for the line to be checked.

The modem won't redial when the line is engaged. The engaged tone just keeps going.

This is easily fixed by setting your modem to return a "busy" result code when an engaged tone is detected, so your terminal software will be aware of it and therefore be able to do something about it. To do this you can use the Hayes command of "ATV1X4" placed in the modem initialisation string.

Once I'm connected to a remote system, I see two of everything I type. So if I type LOGIN, the screen shows LLOOGGIINN.

The duplex settings in your software are wrong. Normally, when you type a character on your keyboard, it gets sent to the remote system and then echoed back. The terminal software displays all received characters on the screen. If the terminal software also displays the characters as you type them, you'll get an extra copy of each character. Toggle your comms software's setting between full and half duplex to fix the problem.

When I type "AT" I don't get the "OK" message displayed on my screen. In fact, I get no responses displayed at all.

You need to tell your modem to display return codes. To do this, use the Hayes command "ATQ0" in the modem initialisation string. You may also need the "ATV1" command to en-

Result string	Result code	Description
OK	0	Command executed OK
CONNECT	1	Connection made
RING	2	Ring signal indicated
NO CARRIER	3	Carrier signal not detected or lost
ERROR	4	Invalid command
NO DIALTONE	6	No dialtone detected
BUSY	7	Engaged signal detected
NO ANSWER	8	No silence detected when dialling a system

Figure 1 - Result codes

sure that result codes are displayed as words, rather than numbers. If you prefer numbers to words then you could issue an "ATV0" command to display all result codes as numbers. The main result codes, with their corresponding result numbers and meanings are shown in Figure 1.

How can I deal with the "call waiting" facility whilst using the modem? Whenever the call waiting tone is received it interrupts the online session and the line drops.

If you have a "call waiting" facility, where you get a soft beep if you have another call waiting to get through, the beeping sound can confuse a modem and cause it to drop the line. But if you set the carrier drop time in your software to more than 1.5 seconds, the problem should be solved. The remote modem won't hang up because it will hear slight attenuation changes or sometimes silence, instead of the beeps that normally cause it to drop carrier.

The other simple solution is just to switch off the call waiting facility if you can easily do so.

I can't get a file transfer in either direction to work. The upload or download doesn't even start.

This is often caused by a mismatch in file transfer protocols. Always ensure that the protocol you are using is also implemented at the other end of the connection. Binary file transfer protocols are required because, simply, some of the characters transmitted in a binary file may be interpreted as end-of-file markers or specific commands to a receiving computer when sent as an ordinary ASCII transfer.

When deciding which protocol to use, you can choose any one that is supported by both ends of the link. Almost all systems and communications packages support Xmodem and Kermit, so you should always find these choices available. More modern systems support Zmodem and, if this is available, you should use it if possible. Zmodem is faster the Xmodem or Kermit, and it also has the ability to let you continue an aborted transfer at the point where the original attempt was abandoned, rather than starting again at the beginning.

The terminal software I'm using keeps telling me that "CTS signal was not detected" and I can't dial anywhere.

This simply means that your communications software isn't communicating with your modem. Check that the modem is actually powered up and that the cable connections to both modem and computer are secured. If the modem is an internal one, make sure it is correctly seated in the slot. Ensure the software is talking to the correct COM port for your modem. This is of great importance when using a portable computer, as most software packages will default to COM 1 whereas a portable computer will almost certainly require you to use COM 2.

RS-232C, also often called just RS-232, is basically a recommendation as to how a serial port should be set to communicate. It has become the industry standard for the serial interface on IBM-compatible PCs, modems, and most other equipment. The pin assignments for RS-232C are as follows, though not all are used.

Pin	Description
1	Protective Ground
2	Transmitted Data
3	Received Data
4	Request to Send
5	Clear to Send
6	Data Set Ready
7	Signal Ground
8	Data Carrier Detect
9	Reserved
10	Reserved
11	Unassigned
12	Secondary Data Carrier Detect
13	Secondary Clear to Send
14	Secondary Transmitted Data
15	Transmit Clock
16	Secondary Received Data
17	Receiver Clock
18	Unassigned
19	Secondary Request to Send
20	Data Terminal Ready
21	Signal Quality Detector
22	Ring Indicator
23	Data Rate Select
24	External Clock
25	Unassigned

Only 9 pins are used in most asynchronous comms applications, out of the 25 pins in the RS-232C standard. Connecting pin 1 is optional in a large number of applications, so the use of only 8 pins is certainly not uncommon. The pins used are: 1, 2, 3, 4, 5, 6, 7, 8 and 20.
Because only these 9 pins are commonly used, most modern PCs use a DB-9 connector rather than a DB-25 one. Just to be really awkward the pin assignments for DB-9 are different to those for DB-25, as shown in the table below:

DB-9 Pin	DB-25 Pin	Description
1	8	Carrier Detect
2	3	Receive Data
3	2	Transmit Data
4	20	Data Terminal Ready
5	7	Signal Ground
6	6	Data Set Ready
7	4	Request to Send
8	5	Clear to Send
9	22	Ring Indicator

Therefore to connect a DB-9 with a DB-25 connector the wiring should be as follows:

DB-9	DB-25
1	8
2	3
3	2
4	20
5	7
6	6
7	4
8	5
9	22

Figure 2 - The RS-232 Interface

PCSA

The Author

Dave Winder is a comms specialist and journalist, and writes the problem page for a variety of communications-related magazines.